7° CONGRESO IAPG Producción y Desarrollo de Reservas
5 - 8 de Noviembre 2019
Mar del Plata, Argentina
INSTITUTO ARGENTINO DEL PETROLEO Y DEL GAS

Implementation and Assessment of Production Optimization in a Steamflood Using Machine-Learning Assisted Modeling

Authors: Pallav Sarma, Tachyus, pallav.sarma@tachyus.com; Carlos Calad, Tachyus, carlos.calad@tachyus.com; Yong Zhao, Tachyus, yong.zhao@tachyus.com

Key Words

Production Optimization, Artificial Intelligence, Machine Learning, Steamflood, EOR, Modeling, Simulation, Mature Fields

Abstract

Data Physics reservoir modeling and optimization was described in detail in as SPE paper (SPE-185507) and can be conceptualized as a physics-based model augmented by machine learning. In brief, the production, injection, temperature, steam quality, completion and other engineering data from an active steamflood are continuously assimilated into the *Data Physics* model using an Ensemble Kalman Filter (EnKF), which is then used to optimize steam injection rates to maximize/minimize multiple objectives such as net present value (NPV), injection cost etc. using large scale evolutionary optimization algorithms. The solutions are low-order and continuous scale, rather than discretized, therefore modeling, forecasting and optimization are significantly faster than traditional simulation.

The goal of steamflood modeling and optimization is to determine the optimal spatial and temporal distribution of steam injection that will maximize future recovery and/or field economics. Accurately modeling thermodynamic and fluid flow mechanisms in the wellbore, reservoir layers, and overburden can be prohibitively resource-intensive for operators who instead often default to simple decline curve analysis and operational rules of thumb. *Data Physics* allows operators to leverage readily-available field data to infer reservoir dynamics from first principles.

This paper presents the results of actual implementation of an optimized steam injection plan based on the *Data Physics* framework. The case study is from a shallow, heavy oil field in the San Joaquin Basin of California, and demonstrates the practical application of *Data Physics* modeling and the ability to explore future injection plans. The model of the field was fit to historical data in June 2017, after which an optimization was performed and a forward-looking production forecast was established associated with a target plan chosen by the operator. This plan was then implemented in the field over the last year. This paper provides a comparison between the field implementation and the model prediction, which allows for model validation and highlights opportunities for further improvement.

For completeness, this paper includes a summary of the modeling and optimization problem and results from the above mentioned paper.

Palabras Claves

Optimización de Producción, Inteligencia Artificial, Machine Learning, Inyección de Vapor, EOR, Simulación, Campos Maduros

Resumen

El modelado y la optimización usando *Data Physics* se describe en detalle en el articulo SPE-185507 y puede conceptualizarse como un modelo basado en física y aumentado por inteligencia artificial. En esencia, la producción, inyección, temperatura, calidad del vapor, datos de completamiento y otros datos de modelado del sistema de inyección de vapor activo se asimilan continuamente en el modelo de Data Physics usando Ensamble de Filtros Kalman (EnKF) el cual se usa luego para optimizar el volumen de vapor a ser inyectado y maximizar/minimizar objetivos múltiples tales como Valor Presente Neto (VPN), costos de inyección, etc. usando

algoritmos evolucionarios de gran escala. Las soluciones son continuas y de bajo orden en lugar de discretas por lo cual el modelado, pronostico y optimización se procesan varios ordenes de magnitud mas rápido que la simulación tradicional.

El objetivo de modelado y optimización de inyección de vapor es determinar el espaciamiento espacial y temporal de la inyección de vapor que maximiza la recuperación y/o la performance económica del campo. El modelado preciso de los factores termodinámicos y los mecanismos de flujo de fluidos en el pozo puede resultar prohibitivo para algunos operadores debido a la intensidad de recursos necesarios, por lo cual los operadores normalmente se inclinan por usar análisis simple de curvas de declinación y/o reglas de oro. Data Physics permite a los operadores apalancar los datos disponibles del campo para inferir la dinámica del reservorio a partir de principios científicos.

Este articulo presenta los resultados de la implementación en el campo de un plan de optimización de inyección de vapor basado en el marco de *Data Physics*. El estudio se realizó en un campo somero de crudo pesado en la cuenca de San Joaquín, Californias y demuestra la aplicación práctica del modelado de *Data Physics* y la posibilidad de explorar planes de inyección futuros. El modelo se ajustó a la data histórica en Junio de 2017, luego se realizó una optimización y se estableció un pronostico de producción a futuro basado en plan de inyección seleccionado por el operador. Este plan fue implementado en el campo durante el ultimo año. El articulo provee una comparación entre la implementación en el campo y las predicciones del modelo sirviendo como validación del mismo; también resalta oportunidades para mejoras adicionales.

Este articulo incluye un sumario del planteamiento del problema de modelado y optimización y su solución del articulo mencionado anteriormente.

Introduction

Steamflooding consists of the process of continuously injecting steam into injector(s) in order to heat the surrounding oil in the reservoir and increase production in nearby producers. The injected steam heats and decreases the viscosity of the oil between the injector and the producing well resulting in greater flow rates and recovery factors [Butler, 1991; Myhill and Stegemeier, 1978; Marx and Langenheim, 1959; Neuman, 1985]. Steamfloods have been very successful in the shallow heavy oil fields of the San Joaquin Valley of California. However, the scale of many such steamfloods with hundreds or even thousands of wells, significant lateral and vertical reservoir heterogeneity, spatial variation of steam quality, varying historical completion quality and methods, different levels of pattern maturity across the fields, etc. also provide significant optimization opportunities. As such, steamflood operation has multiple control variables that can be optimized. Redistribution of steam in existing injectors is one of the most important control variables to optimize in a steamflood to optimally account for the factors mentioned above. A successful redistribution optimization produces a list of time varying injection rate changes for each injector, such as the overall effect of steam on incremental oil production is maximized, and/or steam is optimally cut across already mature patterns, thereby reducing operational costs. Other optimization opportunities like identification and prioritization of infill drilling locations is also important, but will not be a subject of this paper.

For these crucial decisions, operators typically rely on semi-quantitative approaches such as decline curve analysis or simple analytical models. These methods are capable of using only a small portion of available data and function as rules of thumb, resulting in qualitatively optimized reservoir management decisions at best. On the other extreme, some oil companies leverage sophisticated predictive modeling tools. Reservoir simulation is the most advanced technique available today in that it is capable of integrating disparate data sources and predicting over long time horizons [Aziz and Settari, 1979]. While reservoir simulation is an excellent tool for field studies and long-term planning, certain limitations prevent operators from leveraging simulation for day-to-day decision-making.

1. Reliance on geo-modeling workflows requiring months to years of manual setup.
2. Simulation requires hours or days to run a single scenario due to computational complexity.
3. Simulation is impractical to obtain optimal solutions given a limited set of scenarios.
4. Lack of uncertainty quantification because of an insufficient number of models.

5. Sequential data integration results in inconsistencies and failure to honor all the data.
6. Traditional models cannot process the large number of parameters required to increase resolution.
7. It is difficult to integrate data for which physical models are unknown or unclear.

In essence, reservoir simulation enables operators to precisely model the physics of a small number of scenarios; however, it is not designed to produce thousands or millions of scenarios and to automatically update those scenarios based on real-time data. Even for operators with very sophisticated simulators, there is a need for faster models in order to leverage real-time data, achieve prescriptive analytics, and optimize day-to-day operations.

Purely data-driven models lie on the opposite end of the spectrum compared to physics-based reservoir simulation. While simulation models require months to set up and days to run, machine learning models can be built in days and run across a full field in real-time to rapidly explore thousands of scenarios and identify optimal solutions. Although such models offer a significant speed advantage and enable predictive analytics in domains such as unconventionals where the reservoir physics are relatively more difficult to model, the absence of underlying physics prevents machine learning's use for quantitative optimization. The main drawbacks of data-driven approaches include:

1. Models can only accurately predict already measured response in the reservoir.
2. Susceptibility to significant prediction errors due to data quality issues.
3. Inability to accurately predict initial production responses given the absence of data at new locations.
4. Lack of distinction between surface and subsurface causes of production issues.
5. Poor predictivity over longer time horizons and changing reservoir conditions.

Physics-based reservoir simulation and data-driven machine learning offer complementary strengths. An ideal predictive model would combine the speed and flexibility of machine learning with the predictive accuracy of reservoir simulation so that operators could integrate data in real-time to quantitatively optimize reservoir management decisions continuously. Such a model can be used to quantitatively optimize any future performance indicator such as NPV within a "closed-loop reservoir management/optimization" framework [Jansen et al., 2009]. Fig. 1 shows such a closed-loop framework, wherein, the *Data -Physics* models are continuously updated to reflect new data, and the updated models are then used to continuously optimize reservoir management decisions such as steam/water redistribution, infill drilling etc. throughout the life of the reservoir.

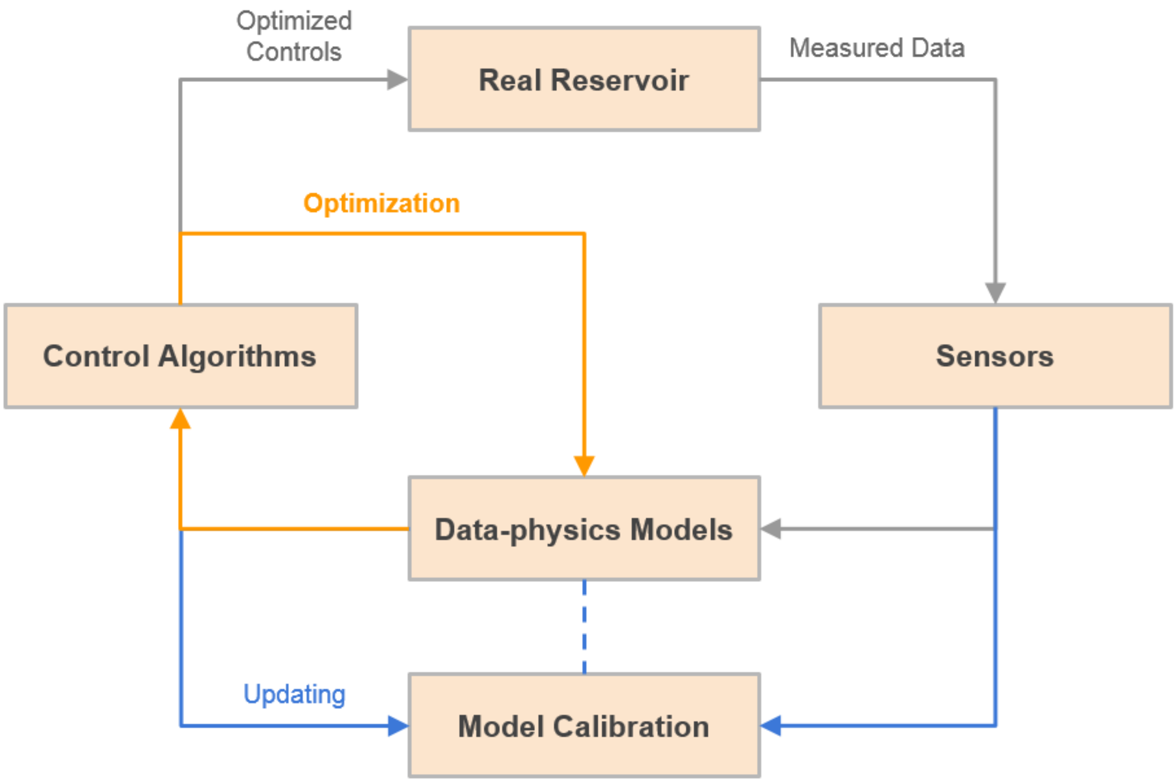


Fig. 1 - Closed-loop reservoir management

Absent fast-running physically accurate models with quantitative production estimates and statistical confidence, such a closed loop process has never before been possible. With a closed loop system where the model is constantly re-run and re-tuned, the reservoir model remains true to actual field conditions at all times and does not fall "out of date."

As such, the primary motivation of *Data Physics* is to create a framework in which petroleum engineers can quantitatively optimize oil production from a reservoir. The goal of quantitative optimization is to predict which actions–for example, on a daily, weekly, or monthly basis–will produce the maximum quantified (predicted and risk-adjusted) return in oil production. Investigating tens of thousands of scenarios that quantitatively simulate specific sets of activities (e.g., increased injection, changes in drilling programs, etc.) allows engineers to select operational plans that meet specific optimization criteria, such as preserving current production rates or optimizing long-term reserves growth. Often, it is necessary or desired to optimize multiple criteria simultaneously. More details on *Data Physics* can be found in SPE-185507 [Sarma et al, 2017].

Case Study

For the field in this case study, production comes from poorly consolidated sands within the Antelope Shale member of the Miocene Monterey formation with porosity averaging 30%, permeability averaging 2,000 mD and net thicknesses typically between 50 and 300 feet. Structural dip is steep at approximately 60 degrees. The reservoirs are shallow, with depths ranging from 200 - 600 feet TVD. Oil gravity is approximately 13° API. Reservoir pressures are well below bubble point and average 50 - 100 PSI. The field has about 200 producers and 30 injectors, producing about 2000bbd of oil and injecting 6500bbd of steam. The field has been under operation for about 8 years, with most of the continuous steamflood starting about 5.5 years ago. The *Data Physics* workflow implemented by the operator in this case study was as follows:

1. Backtest

- o fit historical training data
- o predict historical test data (blind test)
- o statistically compare the forecast with test data to quantify predictivity

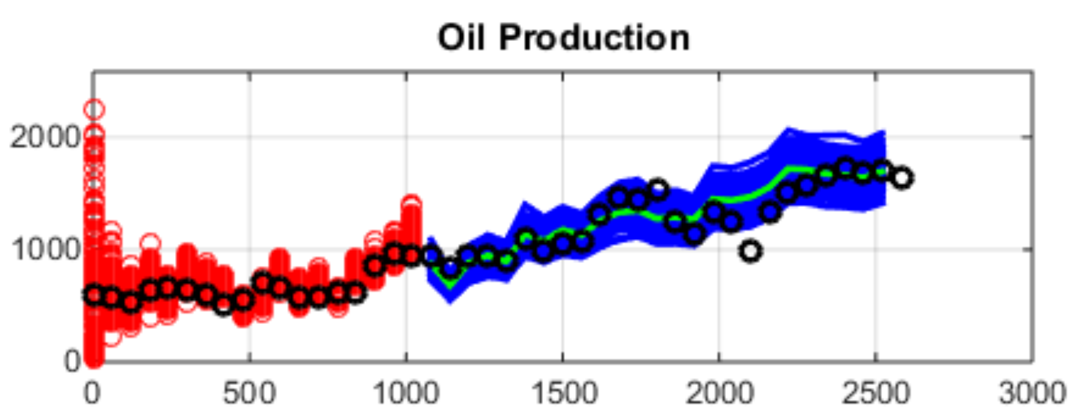
2. Full Fit – train the model on the full historical dataset
3. Optimization – explore future injection distributions
4. Set Target Plan – commit to a target injection plan
5. Implement Target Plan – implement actual steam changes in the field
6. Assess – assess actual performance

The earlier paper explored up to step 3 above, and demonstrated the potential for optimization. This paper extends it further to step 6, and discusses the implementation of an actual target plan and the results observed so far.

Backtest

In the backtest, the historical data was divided into two datasets: training and test. The model was fit using only the training data, and then used to forecast production over the test period, using historical steam injection as the control variable - this allows for verification of the predictive capability of the model. The model was built in June 2017. Note however that due to incompleteness of data during the last three months, the model utilized data only until March 2017. The training period ranged three years, from 2010 to 2013, while the test period ranged four years, from 2013 to 2017. The types of data assimilated are oil and water production rates, steam injection rates, well head pressures at injectors, producer and injector temperatures, completion and location data, and log data such as gamma ray, resistivity, neutron porosity and density porosity.

Figure 2 shows the results of the back-test for field-wide oil production and steam injection. The black circles are the measured data, the red circles are the ensemble members during the fit period, and the blue lines are the predictions from the ensemble. The thick green lines are the mean of the predicted ensemble, and clearly there is a strong agreement between the mean predicted response and the measured data for both oil and steam. Additionally, the uncertainty predicted by the models as represented by the span of the blue lines is reasonable, as the true data is tightly bound by the blue lines. Similar results are also seen for water production and average temperature at the producers. The control variables provided as input for the wells during the prediction period are the well head pressures (WHPs) of the injectors. Further, the producers are considered to be fully open both during fit and prediction periods.



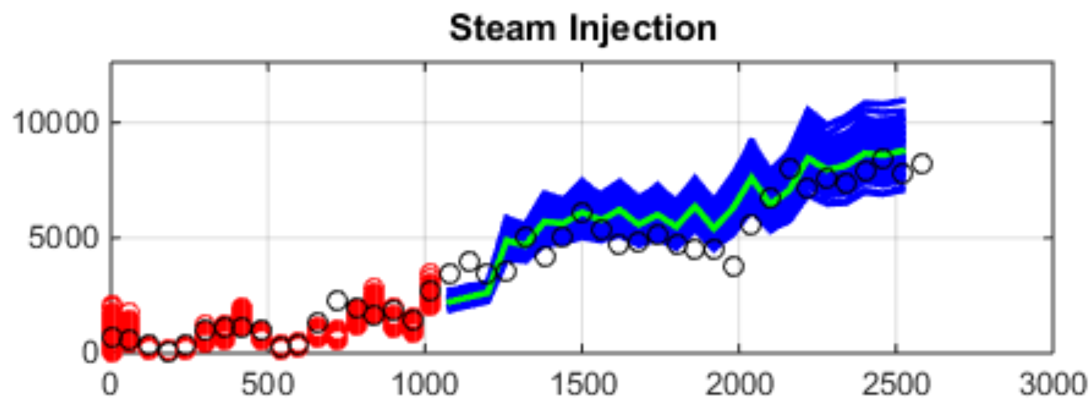


Figure 2 - Comparison of fit and prediction oil production and steam injection to measured data

In order for the model to be useful for optimization, it should be able to predict not only at the field level but also at the well level. However, due to poor quality measured data, back allocation, and model assumptions and errors, it is not possible nor necessary to create a model that can predict perfectly. The left plot of Figure 3 is the scatter plot of actual vs. predicted cumulative oil production during the 4-year prediction period for all wells, and the right plot shows the same for steam injection. The better the prediction, the closer is the dot to the green line. The ability of the model to predict an individual well's production as seen in these scatter plots can be summarized via correlation coefficients such as the Pearson and Spearman (rank) correlations. This is seen in Table 1. We observe a strong correlation of about 0.7 for oil, and slightly poorer correlations of water production, steam injection and temperature. The reason for the lower correlations for steam is due to poor quality WHP data, and this correlation can be improved significantly if WHP measurements are improved. Further, the number of injectors present during the fit period was very low, which could be another cause of the poorer predictions. As for water production, the model is tuned for maximum prediction accuracy for oil production while sacrificing some accuracy for water production. Further, measurement of allocated water production presumably would be of poorer quality. However, if so desired, water rate prediction can be improved by adding additional parameters to the model.

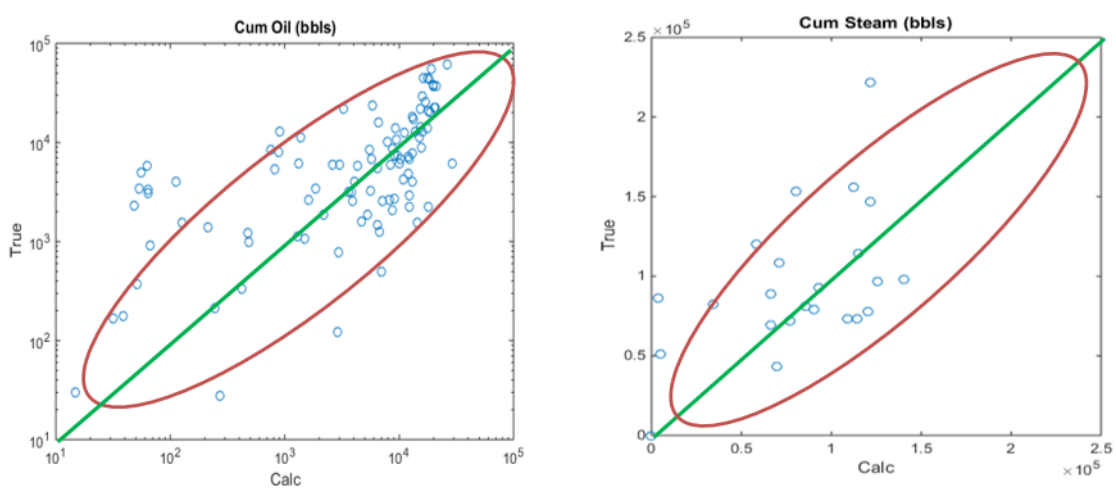


Figure 3 - Scatter plot for calculated vs. measured oil production and steam injection

Table 1 – Backtest Correlation coefficients – modeled vs. measured data				
	Oil	**Water**	**Temperature**	**Steam**
Pearson	0.68	0.43	0.58	0.63
Rank	0.69	0.63	0.62	0.65

Table 1

Additional details on validation of this backtest is provided in SPE-185507 [Sarma et al, 2017]. Given this analysis, it is clear that the model is able to predict future production/injection performance for this field. As such, the model is now fit using all data (training + test data) to further fine tune and improve fit and prediction quality across all wells, and this model will be used for optimization.

Full Fit

Following the validation of model predictivity from the backtest, the fit was extended to include all historical data as shown in Figure 4. This is called a ”full fit”, and in this process the global parameters learned during the back test are not modified, only the local parameters are modified to better fit the latest available data and any new wells that were drilled in the blind test period.

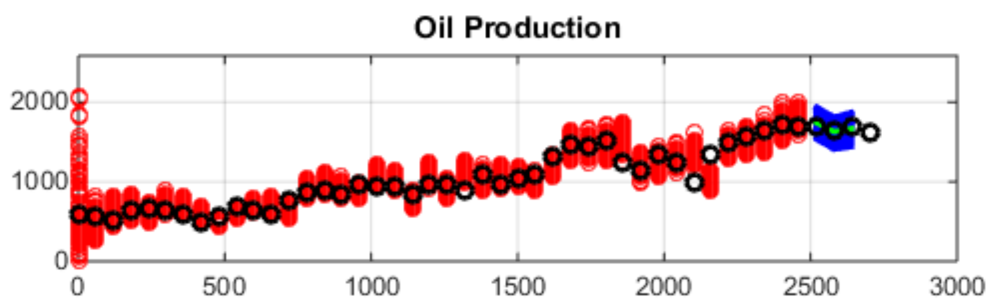


Figure 4 Fieldwide oil production for full fit.

Note that small portion of the data is still used for checking predictivity, and de-biasing the model predictions if necessary. **Error! Reference source not found.** shows the correlations for the restricted prediction period. As expected, the correlations are better as the prediction period is shorter. Steam correlations are 1 as the model is switched to rate control.

Table 2 – Full Fit Correlation Coefficients – modeled vs. measured data				
	Oil	**Water**	**Temperature**	**Steam**
Pearson	0.74	0.59	0.77	1.00
Rank	0.72	0.67	0.75	1.00

Optimization

In SPE-185507, a single objective NPV optimization was performed, with steam capacity constrained to be the same as current. The optimal plan from that optimization was analyzed using various statistical methods and reservoir engineering principles for validation. Once the model and optimization was validated, a more detailed multi-objective optimization was performed for the purpose of implementation, which is discussed here.

After a full fit was obtained, production forecasts for thousands of injection plans were explored using a Metamodel-Assisted Evolutionary Algorithm (MAEA) [Kyriacou, S., 2013]. This method efficiently searches the solution space of possible future injection scenarios to determine an "efficient frontier" of optimal plans. In this case, multiple objectives were defined: maximize production over 15 years, maximize NPV over 15 years assuming a 10% discount rate and oil/steam/lift price projections, and minimize steam injection. Figure 5 shows dozens of injection plans; each point on the chart represents a time-varying list of injection rates for every injection well.

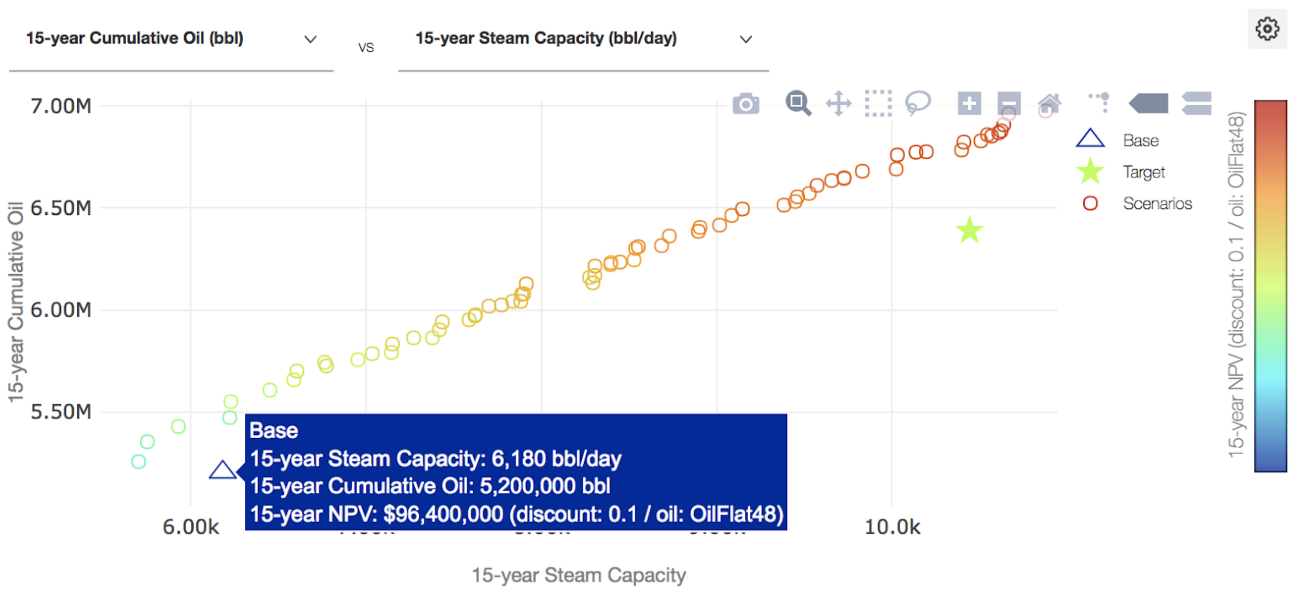


Figure 5 The "efficient frontier" of future injection plans. The triangle represents the current operating plan at the time of modeling, termed the base case. Each circle represents a particular model-predicted optimal injection plan.

Target Plan

The star in Figure 5 represents a particular scenario that was modified by the operating company to comply with facilities constraints. The modification caused the production forecast to be slightly below the efficient frontier, but still remain well-above the base case. This is the scenario that was chosen as the target operating scenario. The injection plan is shown aerially in Figure 6 wherein each bubble corresponds to an injection well. Bubble size corresponds to the cumulative injection delta between the current plan and the target plan, while color indicates an increase (green) or decrease (red). Based on the model and optimization plan, the operator essentially planned to increase their steam injection rate from ~6,200 bbd to ~10,400 bbd field wide.

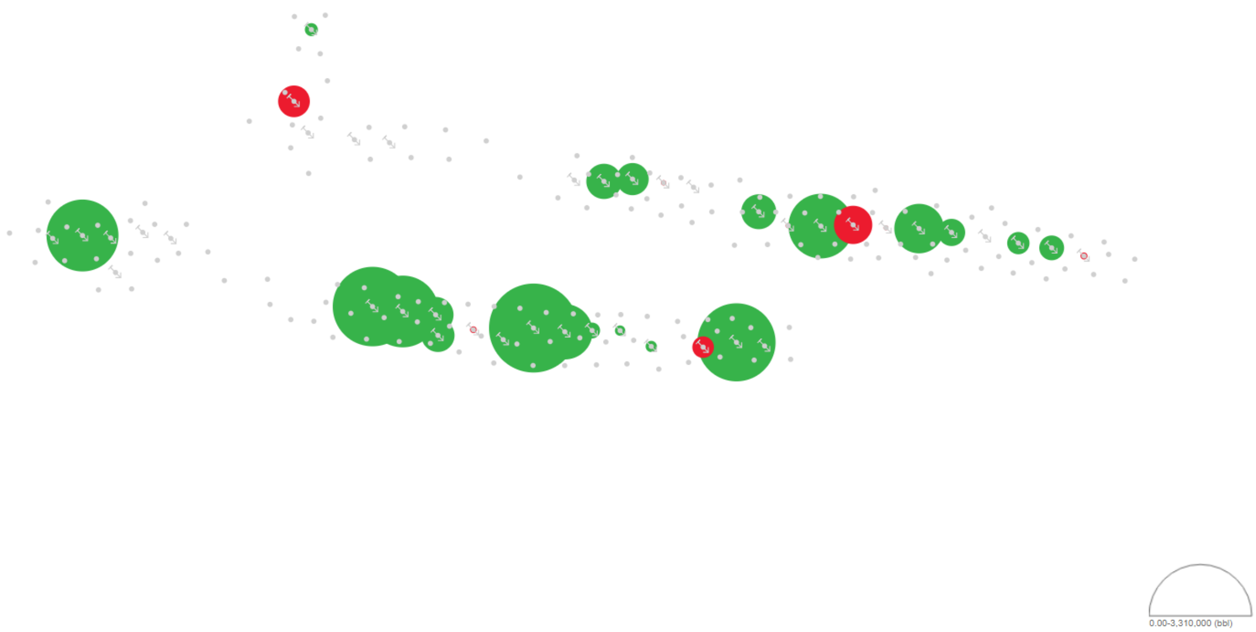


Figure 6 - Target injection plan bubble plot.

Figure 7 shows the difference between base plan and target plan. There is an increase in steam injection of 4,200bb, and an expected increase in oil production of 1.2MM bbl. over the next 15 years if the plan is followed as-is.

15-year	Base	Target	Delta
Steam Capacity (bbl/day)	6,200	10,400	4,200
Cumulative Oil (bbl)	5.2E+06	6.4E+06	1.2E+06
NPV (USD)	9.6E+07	1.05E+08	8.6E+06

Figure 7 - Base vs target plan

Implementation and Assessment

A year after fitting the model, optimizing, and selecting the target plan, conformance of actual injection to target plan and model predictivity were assessed. Figure 8 shows conformance to the target plan over last year. It is clear that while the operator injected significantly more steam compared to the base case, it still fell short of the target plan due to various operational reasons. It takes time to make major operational changes such as the significant change in injection from base to target, which is why we see a slow ramp up in injection from Jul 2017, when the implementation of the target plan started. In any case, such deviations from planned operation are to be expected, which provides further support for continuous closed-loop optimization, as constraints and plans change over time.

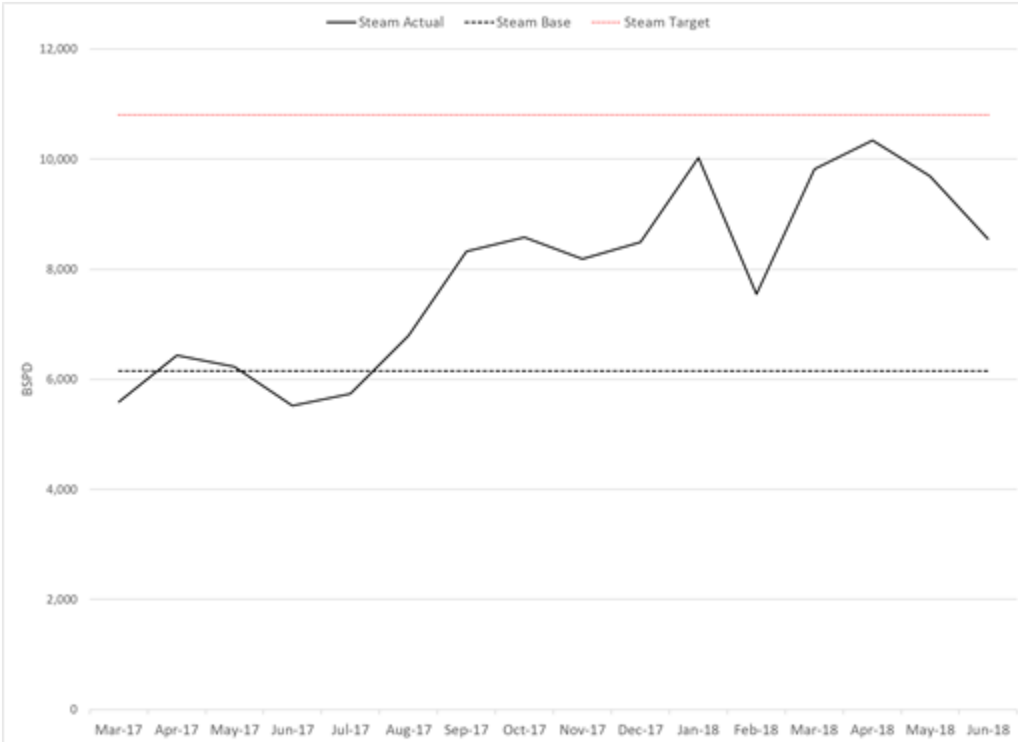


Figure 8- Conformance to target plan

Since the target plan was not implemented exactly, we cannot use the target production forecast as a benchmark for actual production. Nevertheless, model predictivity can still be tested by inputting actual injection rates into the existing model, along with other operational constraints and activity such as cyclic injection over the last year. Like a backtest, this is a true blind test, as this new data was not available when the model was built in June 2017. Figure 9 compares the P50 model prediction vs. true production. We observe that the model response is quite close to true production, and increases and decreases in the production trend are captured well, but in general the model is under-predicting slightly. While it is a positive for the operator to realize more production than predicted, it is useful to analyze why the model predicted expected production was too low.

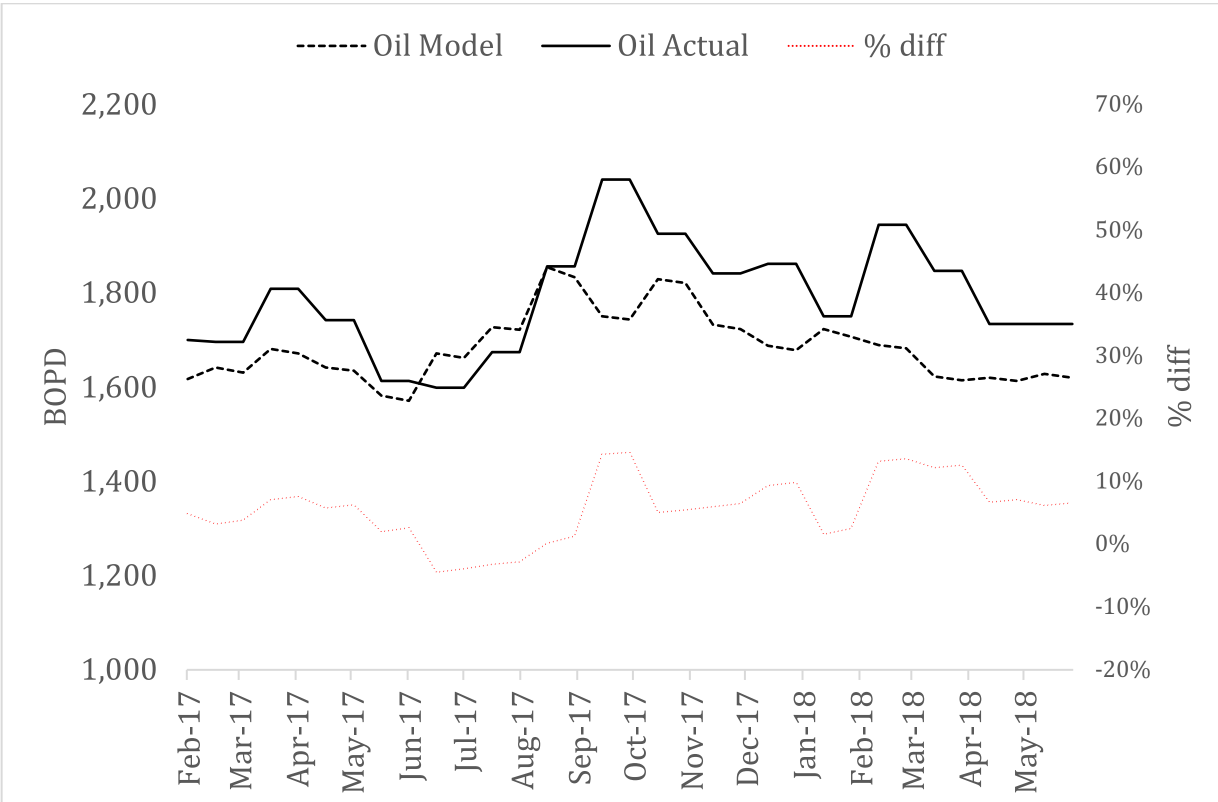


Figure 9 - Comparison of expected model forecast with actual production.

To analyze this further, Figure 10 shows the P10-50-90 model predictions in blue and the true production in black. Production exceeded the p90 prediction in a few time steps, but overall the true production usually falls between the P10-P90 intervals. It is clear, however, that the P50 model tends to under predict, particularly later in time. The underlying cause of this was found to be the over/under burden heat loss model, which predicted greater heat loss than what actually happened, which caused predicted reservoir temperature to be lower, leading to a low production forecast.

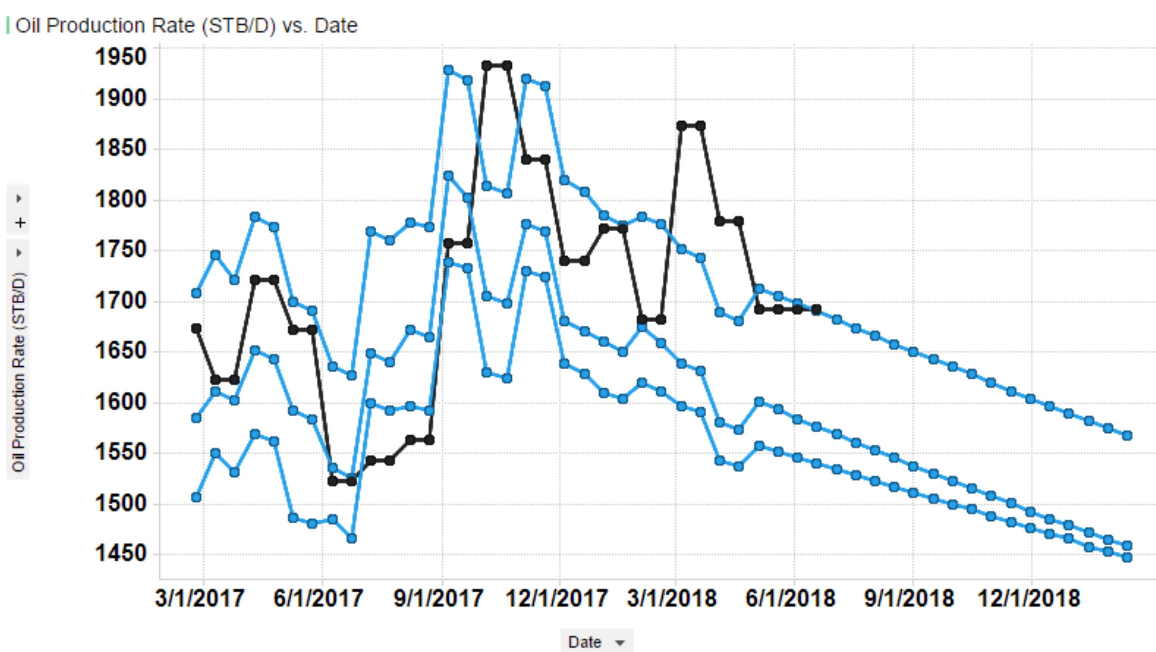


Figure 10 - Comparison of P10-50-90 model-prediction (blue) with actual oil production (black).

Figure 11 shows the four main subdivisions or leases of the field and Figure 12 shows the model predictions and true production broken out accordingly. We observe that the field level prediction is generally more accurate than the lease level prediction. This is likely related to errors in allocation of production from field-level measurements to well and lease. Field-level production data is generally very accurate due to sales measurements, but the process of back allocating that data to wells and leases is prone to error due to poor quality well test data.

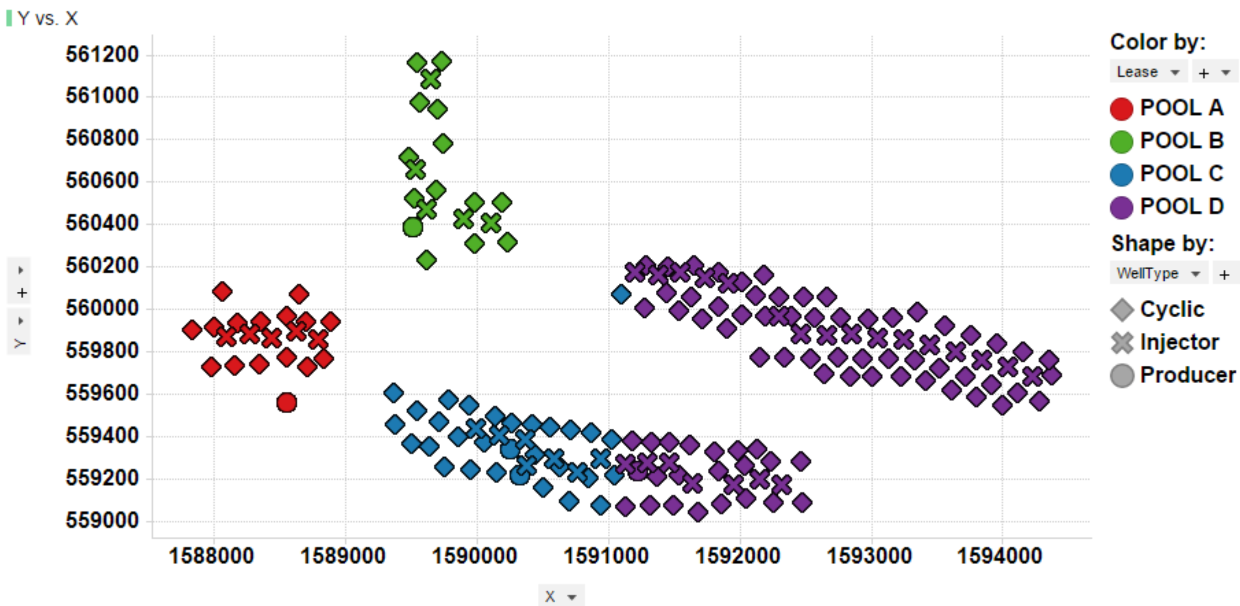


Figure 11 - Four main leases.

Additionally, due to the statistical nature of the modeling process, results are usually better when more wells are aggregated together for analysis, as errors tend to cancel out over larger sample sizes. In other words, it is difficult to predict well the behavior of a single well due to the significant noise in well-level data, but predictions usually improve as data is aggregated over a large number of wells. To demonstrate this point, the largest lease, named Pool D, shows the most accurate predictions, with the true data falling within the P10-P90 range more reliably than the smaller leases. Predictions for Pool B follow the trends quite well, though the predicted range is too narrow. Predictions for Pool A tracks well for the first six months, after which the model under predicts. For Pool C, the model over-predicts throughout.

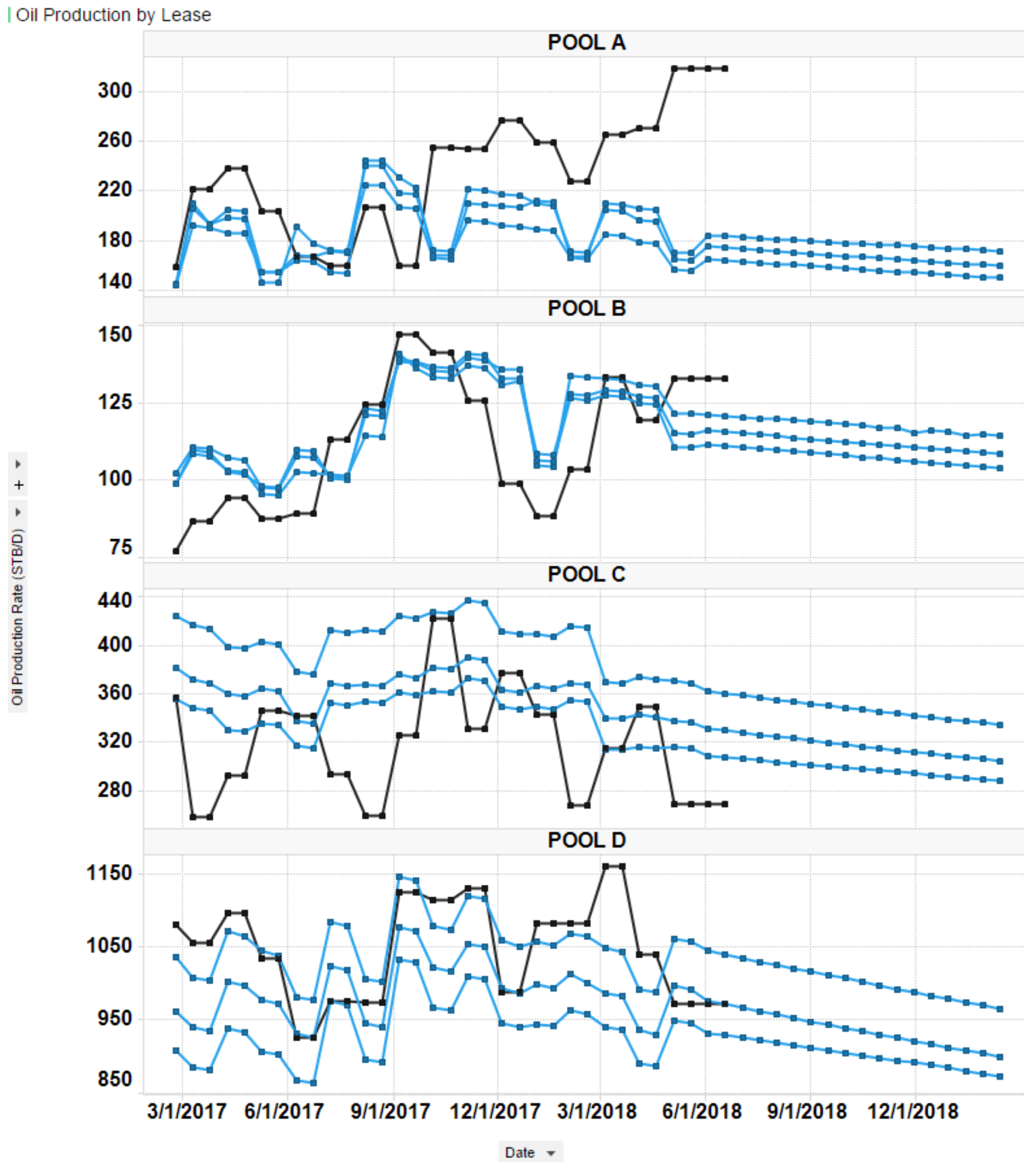


Figure 12 - Model predictivity by lease

Conclusion

The *Data Physics* modeling technique was described and its practical application in the context of a heavy oil steamflood was outlined. A model from one year ago was the basis of a significant change in field development strategy and was used to benchmark actual performance over the intervening time period. Field wide predictivity was within confidence intervals to support further reliance on the model for future operating decisions. Overall, this case study demonstrates that the prescriptive modeling capability of *Data Physics* is well-suited to perform such analysis and provide a fast, actionable forecast that would not be available by other means.

The reliability of the model is notable because it served not simply as an academic exercise but to support a 60% injection increase from the base case, representing a major operational change in the field, as seen in Figure 8, particularly given the significant extrapolation involved in increasing injection by over 60% from the base case. The ability to extrapolate beyond observed data is one of the key benefits of *Data Physics* over simple machine learning approaches. The results in Figure 5, which were calculated using a model built over a year ago, accurately predicted significant upside from a nearly twofold injection increase. The resulting operational change corresponded to significant incremental production beyond the expected base case decline.

References

Aziz, K. and Settari, A., 1979. *Petroleum Reservoir Simulation.* Applied Science Publishers, London, 135-139

Butler, R., 1991. *Thermal Recovery of Oil and Bitumen*. Prentice-Hall.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics.* November 2003, Volume 53, Issue 4, pp 343–367.

Jansen, J.-D., Brouwer, R., and Douma, S. G., 2009. Closed Loop Reservoir Management. Society of Petroleum Engineers. doi:10.2118/119098-MS

Kyriacou, Stylianos A. 2013. "Evolutionary Algorithm-based Design-Optimization, Methods in Turbomachinery". PhD Thesis, NTUA.

Marx, J.W., Langenheim, R.H., 1959. Reservoir Heating by Hot Fluid Injection. In Trans., AIME, 216, 312–314.

Neuman, C.H., 1985. A Gravity Override Model of Steamdrive. *J Pet Technol*, 37 (1): 163-169.

Sarma, P, Kyriacou, S., Henning, M., Orland, P., Thakur, G., Sloss, D. 2017. Redistribution of Steam Injection in Heavy Oil Reservoir Management to Improve EOR Economics, Powered by a Unique Integration of Reservoir Physics and Machine Learning. Paper SPE 185507-MS presented at the SPE Latin America and Caribbean Petroleum Engineering Conference in Buenos Aires.

Myhill, N. A., and G. L. Stegemeier. "Steam-drive correlation and prediction." *Journal of Petroleum Technology* 30.02 (1978): 173-182.